Evaluating The Effect of County Diversity on Educational, Economic, and Health Outcomes

Kenny Groszman, Jack Petersen, Bing Xu

Abstract:

It is often claimed that diversity brings with it many benefits. One could envision that growing up in a diverse community with exposure to many different cultures and perspectives could expand one's own understanding of the world and better-prepare someone for living in an increasingly diverse country and society. It is our claim hypothesis that the diversity of one's community has wider-reaching and quantitatively-measurable downstream effects on that community's educational quality, economic success, and health.

In order to determine whether this is true, we first need a measure of diversity for a geographic area. Pulling household-level US census data from IPUMS (Integrated Public Use Microdata Series), we are able to come up with a measure of diversity, which we dub the Diversity Index, with county-level granularity. We can then analyze relationships between diversity and demographic trends (educational, economic, and health) across the different counties in the IPUMS dataset. Through this, we hope to establish the value of growing up in a diverse community.

Progress

For this project update, we have chosen to first attack the effects of diversity on educational outcomes. In order to do this, we needed to do the following:

- Create Diversity Index
  - Group individual records from IPUMS by county
  - Establish Diversity Index that adequately ranks counties based on the heterogeneity of their populations
- Assess effect of diversity on educational outcomes
  - Pull and clean educational data from IPUMS and other sources so that it provides a summary of educational quality per county
  - Analyze relationships, or lack thereof, between diversity and educational outcomes

Creating Diversity Index

Formatting and Cleaning IPUMS Data

Data obtained from IPUMS contains geographic information about the household. Specifically, it contains State-level and County-level Federal Information Processing (FIP) codes. These serve as unique identifiers of counties, and are the unique keys for the county-level database that we are building. Using the *dplyr* and *stringr* package, we were readily able to organize this data by county and to format the IPUMS geographic information into a form compatible with FIPS. This allows us to easily compare this data to county-level educational, economic, and health measures.

Creating Diversity Index

Upon summarizing race data by county, race information is given as the proportion of the county population that falls into 8 different groups: * White * Black * American Indian/Alaskan Native * Chinese * Japanese * Other Asian / Pacific Islander * Othe Race * 2+ races

We needed a measure of Diversity that follows the following conditions: 1. The county with the highest possible Diversity Index Score is one with the population spread evenly across the 8 groups (i.e. 12.5% in each group) 2. A county with the lowest possible Diversity Index Score is one with the population entirely in one of the groups (i.e. 100% White, 0% everything else)

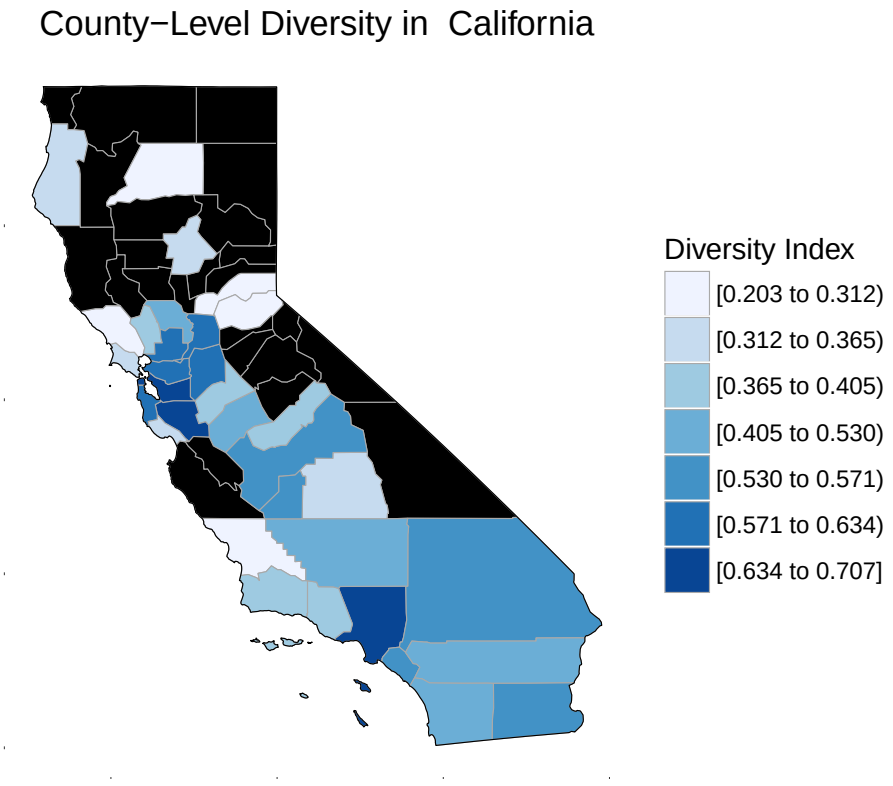
A formula that meets this criteria is for county i is as follows, for races $R \in \{White, Black, ..., 2 + Races\}$:

$$DiversityIndex_i = 1 - \sum_{j \in R} x_{ij}^2$$

where x_{ij} is the proportion of race j in county i

Results

The Diversity Indices of the 521 counties in the IPUMS dataset range from .054 to .778, and can me mapped in a choropleth map on the USA. Since our data does not include all of the counties in the US, a nationwide map is quite spares. However, as an example, here is a choropleth map of the counties in California, where black counties correspond to counties with no data available.



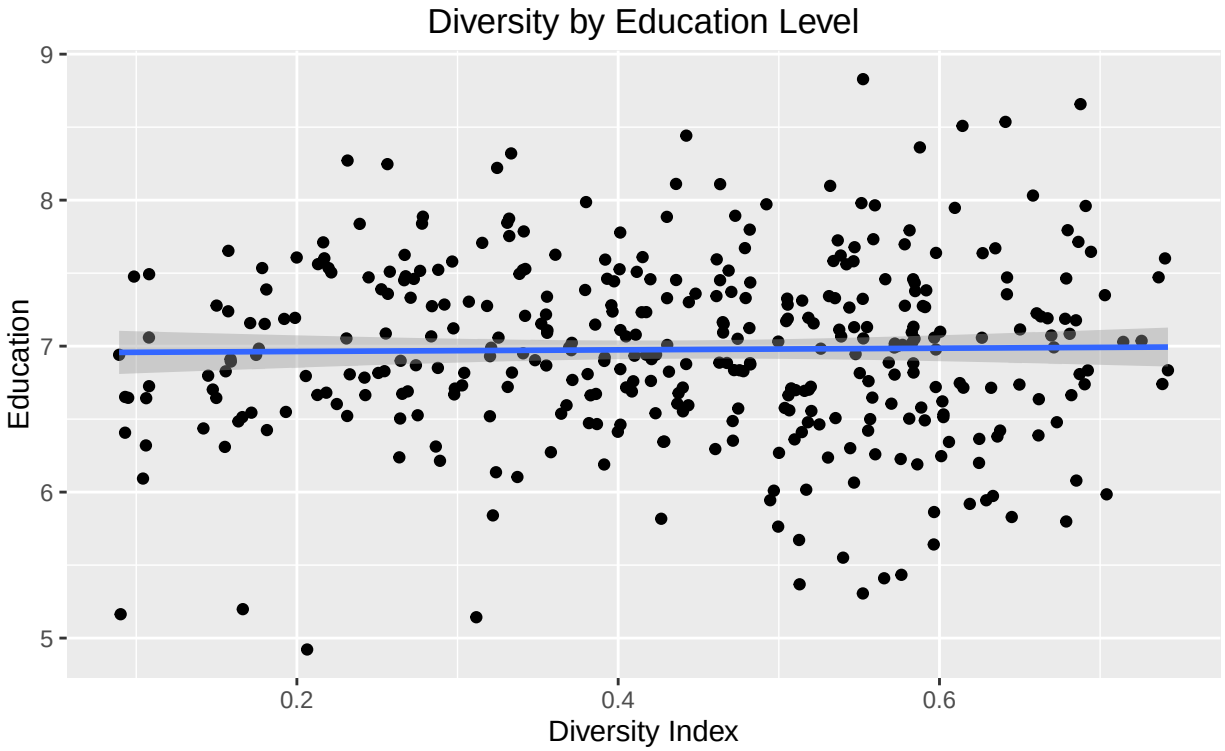
Analyzing Diversity Effects on Educational Outcomes

Here, we pull educational data frmo various sources and compare it to county-level diversity data. The hops is that we would see a relationship between diversity levels and the educational trends in a county. We wish to look at the following measures for educational outcomes:

- IPUMS Educational Attainment measure, which is the highest level of school or degree completed, averaged per county
- Percentage of Adults with:
  - No High School diploma
  - High School diploma
  - Some college or Associate's Degree
  - A completed Bachelor's Degree

1. IPUMS Educational Attainment

IPUMS Educational Attainment measures the highest level of school or degree completed. We averaged this measure over each county and performed a regression to see whether there is any correlation between county diversity and educational attainment (as measured by IPUMS)



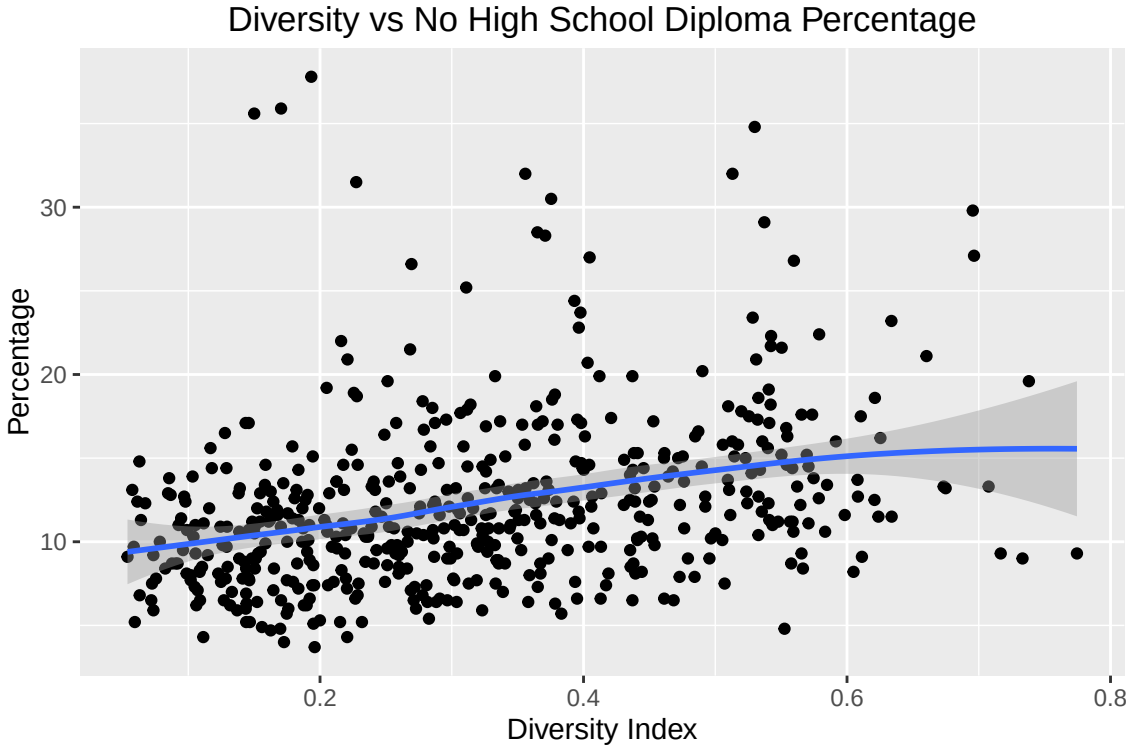
Commentary and Interpretation:

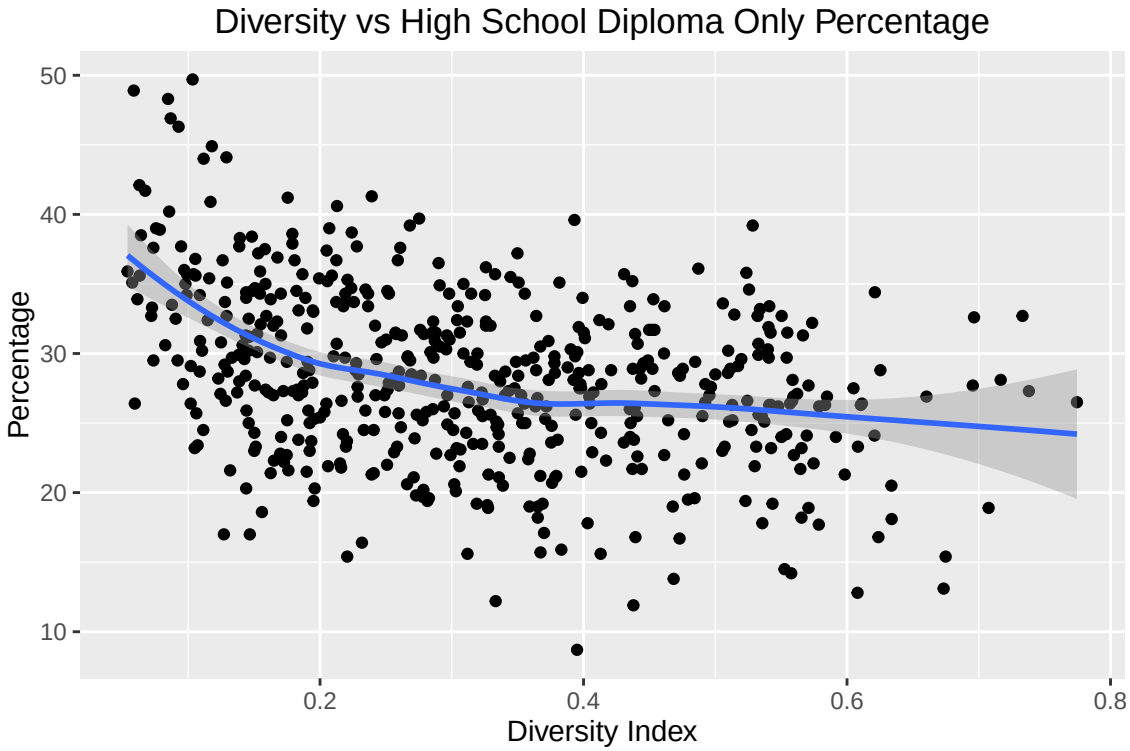
We see that diversity level in a county is *not correlated* with the education level of the county. In our graphic, the trend line we appended is nearly flat, suggesting that the metric we used to measure education is independent of diversity index. Moving forward, we will consider other measures of education, combining different factors given to us to move past the single variable that we used here.

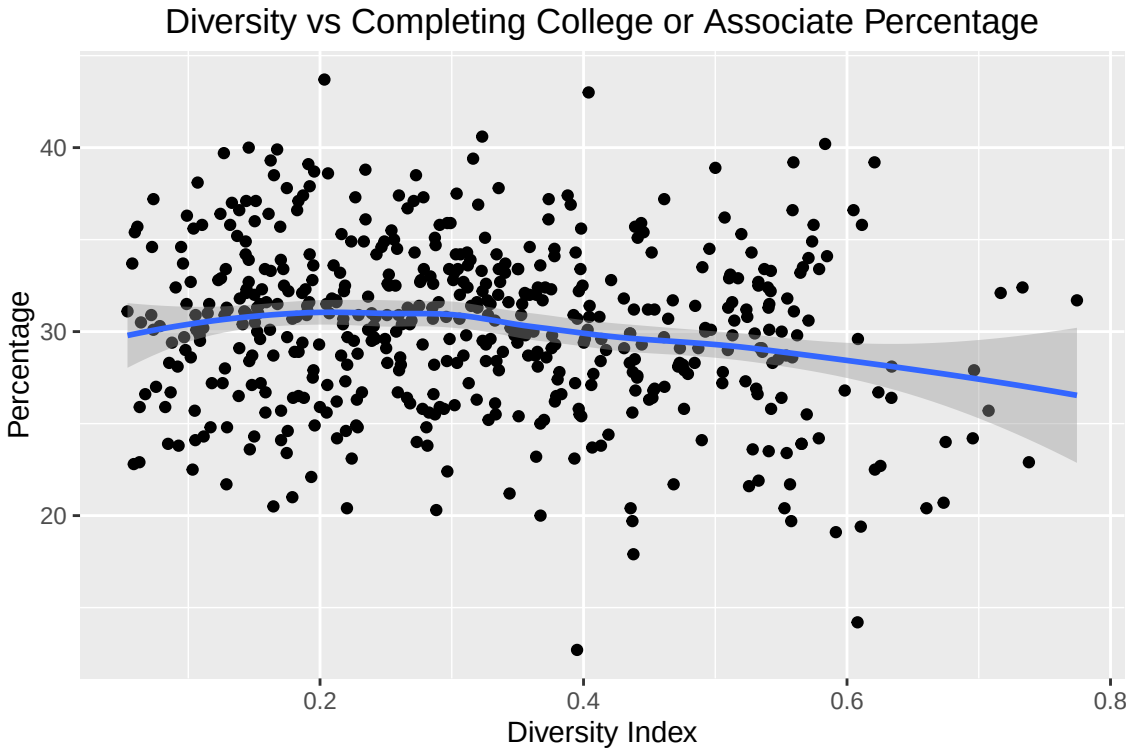
2. Relationship between Diversity and Percentage of Adults with Particular Levels of Educational Attainment

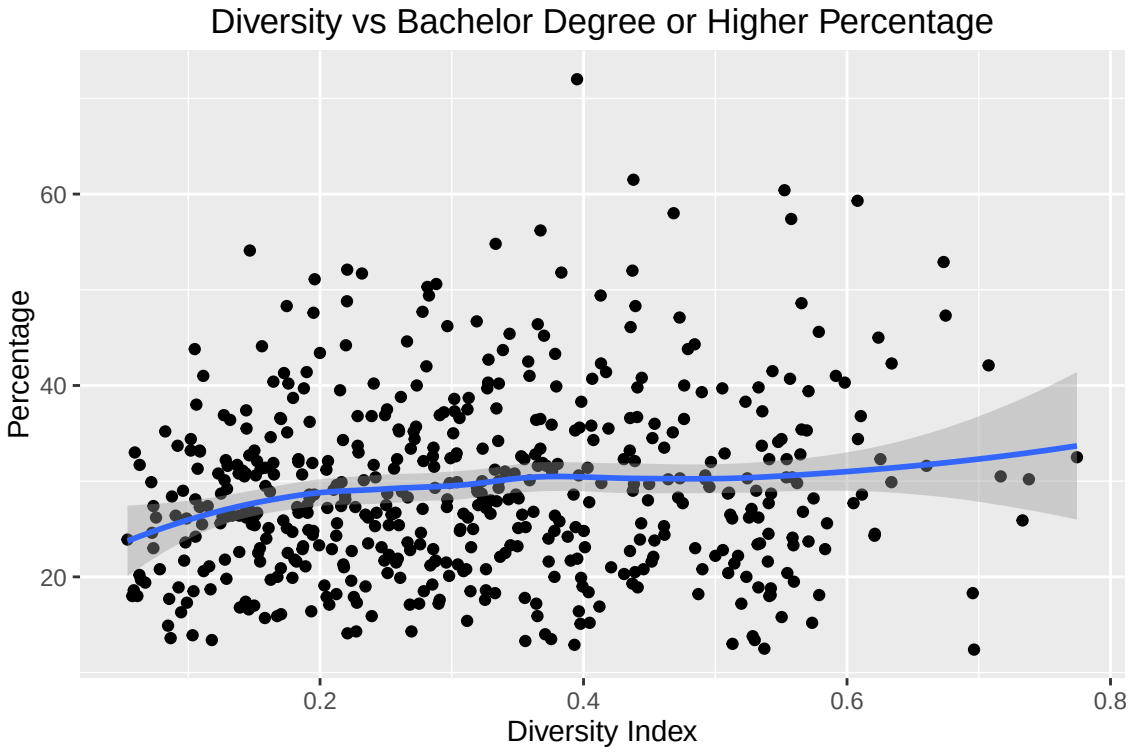
From the US Department of Agriculture Economic Research Service, we were able to find a data set that includes the percentage of adults attaining certain education levels for each county. So we are interested in finding whether there are relationships bewtween the diversity we devised and percentages of completing degrees. There are four percentages, which are percent of adults with less than a high school diploma, percent

of adults with a high school diploma only, percent of adults completing some college or associate's degree, percent of adults with a bachelor's degree or higher.









Commentary and Interpretation:

From graphs above we can see that all rates seems not to have strong correlation with the diversity index. We also conducted the linear regression analysis and all regressions have a relatively poor R^2, with the maxumum at 13.27% (the regression between diversity index and the percent of completing a high school diploma only).

We can find that there is a decreasing trend between the diversity index and the percentage of adults completing a high school degree only. We can find that counties whose high school diploma only percentage is high are having a relatively lower diversity index, which means that the race are more concentrated in these counties.

As for the relationship between the percentage of less than a high school diploma and the diversity index, the trend is a bit opposite. Highly diversified counties are a bit likely to have a higher percentage. There are some outliers and these outliers implies that counties with either low or high diversity indexes could have a high percentage.

While for the last two graphs, the relationships between variables are quite weak and there are many outliers in degree percentages.

Next Steps

In all, these results show weak correlations between county diversity and educational attainment. We hope that examining other damographic outcomes, such as economic outcomes (average wages, wage gaps, number of new businesses per year and types of businesses) and health trends (incidences of obesity, AIDS, diabetes, etc.) will show better results. Even if they don't we will be able to conclude that there is not a huge effect of diversity on these measures, which is valuable to know as well.